BBL536E Homework-1 Report

Problem-1

Task1:

- I found optimal alpha parameters of Ridge model in parameter set {0.001,0.01,0.1, 1.0, 10.0} for y1 output as 0.001 and for y2 output as 0.001.

Task2:

- Best parameter set for y1 is:
  {'max_depth': 250, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}
- Best parameter set for y2 is:
  {'max_depth': 250, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 10}

For task1 and task2, table of mean and standard deviation of 'Mean Squared Errors' and 'Mean Absolute Errors' is:

	Mean Absolute Error		Mean Squared Error	
Output	Random Forest	Ridge Regression	Random Forest	Ridge Regression
Y1	0.31±0.03	2.091±0.23	0.21±0.06	8.71±1.74
Y2	1.01±0.16	2.26±0.26	2.87±0.82	10.35±2.53

Screenschot of output of the code is as follows:

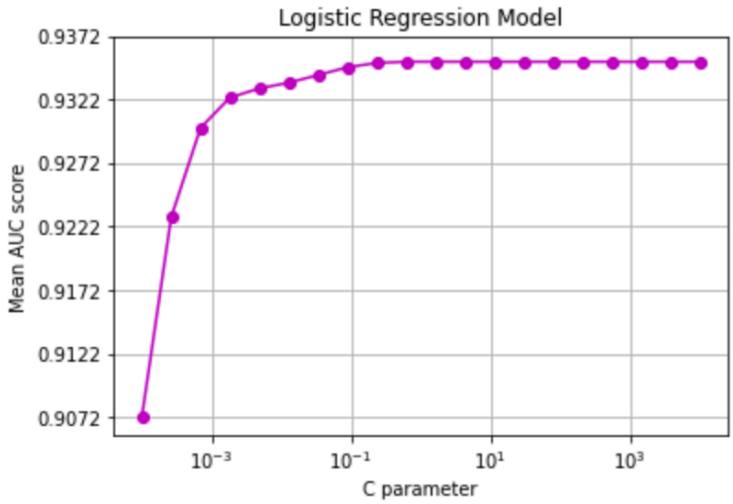
```
problem-1 Task-1
Optimal alpha parameter of Ridge model for target y1:  0.001
Ridge mean score of mae value for Y1 is:  2.0910518036913537
Ridge standard deviation of mae value for Y1 is:  0.2361482238285338
Ridge mean score of mse value for Y1 is:  8.710823249413732
Ridge standard deviation of mse value for Y1 is:  1.7413434482072414

Optimal alpha parameter of Ridge model for target y2 is:  0.001
Ridge mean score of mae value for Y2 is:  2.2667374490882337
Ridge standard deviation of mae value for Y2 is:  0.26950209601233466
Ridge mean score of mse value for Y2 is:  10.353927888203657
Ridge standard deviation of mse value for Y2 is:  2.531010179953866


Problem-1 Task-2
Best parameters set for target y1:
{'max_depth': 250, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}
Forest mean score of mae value for Y1 is:  0.31243733096377324
Forest standard deviation of mae value for Y1 is:  0.03968503541248176
Forest mean score of mse value for Y1 is:  0.21742151000691998
Forest standard deviation of mse value for Y1 is:  0.0625389148787975

Best parameters set for target y2:
{'max_depth': 250, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 10}
Forest mean score of mae value for Y2 is:  1.017473998632946
Forest standard deviation of mae value for Y2 is:  0.1654416445342916
Forest mean score of mse value for Y2 is:  2.872290699128504
Forest standard deviation of mse value for Y2 is:  0.8296570784490276
```

Problem-2

Task1:

In this task, for built Logistic Regression model**,** output of code for **best C parameter with highest AUC score** and **mean AUC score vs C parameter graph** for 5-Fold, 5 repetition model is as follows;

```
best C parameter is {'C': 1.623776739188721} with 0.9351636957716422 AUG score
```



Task2:

For built Random Forest model, output of **best parameter set with best AUC score** for grid search with 3-Fold, 3 repetition cross-validation strategy is as follows;

```
Best parameter set with  0.9448275462821978 AUC score is:
{'max_depth': 150, 'min_samples_leaf': 3, 'min_samples_split': 3, 'n_estimators': 1000}
```

Task3:

For built Neural Network model, output **of best parameter set with best AUC score** for grid search with 3-fold, 3 repetition cross-validation strategy is as follows;

```
Best parameter set with : 0.9361001322834065 AUC score is:
{'alpha': 0.1, 'hidden_layer_sizes': (10, 10, 10)}
```

Task4:

Classification reports for Logistic Regression, Neural Network and Random Forest models are as follows;

(For parameter of Logistic Regression, C=1 is used, For Neural Network model and Random Forest model, output parameter sets found in task2 and task3 are used.)

Classification report for Logistic Regression:

	precision	recall	f1-score	support
0	0.93	0.97	0.95	365480
1	0.66	0.42	0.52	46400
accuracy			0.91	411880
macro avg	0.80	0.70	0.73	411880
weighted avg	0.90	0.91	0.90	411880

Classification report for Neural Network

	precision	recall	f1-score	support
0	0.95	0.96	0.95	365480
1	0.62	0.56	0.59	46400
accuracy			0.91	411880
macro avg	0.78	0.76	0.77	411880
weighted avg	0.91	0.91	0.91	411880

Classification report for Random Forest

	precision	recall	f1-score	support
0	0.93	0.97	0.95	365480
1	0.68	0.46	0.55	46400
accuracy			0.91	411880
macro avg	0.81	0.72	0.75	411880
weighted avg	0.91	0.91	0.91	411880